MASSACHUSETTS INSTITUTE OF TECHNOLOGY
A.I. LABORATORY

Artificial Intelligence
Memo No. 430

June 1977

PLAIN TALK ABOUT NEURODEVELOPMENTAL EPISTEMOLOGY

MARVIN MINSKY

ABSTRACT

This paper is based on a theory being developed in collaboration with Seymour Papert in which we view the mind as an organized society of intercommunicating "agents". Each such agent is, by itself, very simple. The subject of this paper is how that simplicity affects communication between different parts of a single mind and, indirectly, how it may affect inter-personal communications.

This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory's artificial intelligence research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014-75-C-0643.

This paper will be presented at the Fifth International Joint Conference on Artificial Intelligence, Cambridge, Massachusetts, August 1977.

This paper is based on a theory being developed in collaboration with Seymour Papert [Note 1] in which we view the mind as an organized society of intercommunicating "agents". Each such agent is, by itself, very simple. The subject of this paper is how that simplicity affects communication between different parts of a single mind and, indirectly, how it may affect inter-personal communications.

To set the stage, imagine a child playing with blocks, and think of this mind as a society of interacting agents. The child's principal surface goal at a certain moment might emerge from an active WRECKER:

> WRECKER wants to push over a tower, to see and hear it crash.

WRECKER devises a plan that requires another agent, BUILDER, to make a tower, which will later be toppled.

> BUILDER wants to build the blocks into a high tower.

BUILDER's first steps yield a tower that is not high enough (in the view of some Critic set up by WRECKER). The response to this criticism is to add another block to the tower.

> BUILDER must call on another agent, PUT, who knows how to move a block to a specified location. And PUT will have to call an agent to GRASP. Both PUT and GRASP will have to call on TRAJECTORY specialists for moving HAND.
>
> There is a potential conflict between BUILDER and WRECKER. BUILDER wants to persist in making the tower higher, while WRECKER is satisfied with its height and wants to complete his plan of knocking it over. The conflict becomes a problem for a superior PLAY-WITH-BLOCKS agent who started the activity; both BUILDER and WRECKER are competitors for his favor.
>
> The dispute might be settled locally at the level of PLAY-WITH-BLOCKS, but there is another problem. The internal conflict might weaken the status of PLAY-WITH-BLOCKS, who himself is only a minion of an even higher-level agent, PLAY, who (in turn) is already engaged in a conflict with the powerful
>
> I'M-GETTING-HUNGRY. If the latter takes control, the structure that PLAY has built will start to disintegrate -- not the tower of blocks, but the society

of agents organized to build it! Even so, probably WRECKER would win a small victory in the end (even as he fades away), when the child smashes the tower on his way out. [Note 2]

It is not our purpose here further to discuss conflict and control in the Society of Minds, but only communication. If each of these agents were real, separate people, e.g., a group of children, then it would be reasonable for BUILDER to use a natural language to say to PUT like:

Put the Green Block on top of the Tower.

But, if each agent is only a small component of a single mind, he cannot use anything like a natural language because:

Each agent would need syntactic generation and analysis facilities. Our agents are just intelligent enough to accomplish their own specialized purposes, so this would be an enormous burden. [Note 3]

For agents to use symbols that others understand, they would need a body of conventions. We want to bypass the need for dispersed but consistent symbol definitions.

Even conventions about ordering of message elements would be a burden. This might be no problem in a serial computer, but we are concerned here also about how a brain might work as a parallel computer.

In fact, we don't think the agents should use a language at all -- that is, an ordered string of symbols. We will propose a parallel, spatial scheme. First we ask: what does the recipient of such a message really need to know? In the case of BUILDER's message to PUT

TOWER-TOP--------trajectory destination
GREEN BLOCK----------trajectory origin
HAND--------------instrument of action

are all PUT has to know to get started, although its sub-specialists will have to know more. Thus, when PUT calls upon GRASP, the latter may need to know the size, shape, and weight of GREEN BLOCK. GRASP and PUT will need the locations of GREEN BLOCK and of TOWER-TOP. But none of these additional items need be in the surface message from BUILDER since they would only have to be passed along to sub-specialists.

Such specification-lists are familiar under such names as

---"attribute-value list" (AI)
---"frame-terminals" (AI)
---"calling sequence" (programming)
---"case slots" (linguistics)

In computer programs, one does not usually transmit the actual values of arguments to subroutines, especially when they are complex. Instead, one transmits only "pointers" -- symbols designating the memory-locations of the data. In our inter-agent situations, the data *is* usually complex, because each item is an entry to a semantic network or else is not yet completely specified.

In fact, we shall argue for a system that does not even transmit pointers! To put the proposal in perspective, we list a few alternative schemes:

1. Send a list of attribute-value pairs. The recipient has to decode the symbols and assign the values.

2. Send an ordered list of values. The recipient must know where to put each item.

3. Send an ordered list of pointers. The recipient must understand the ordering and the address code.

4. Send a linear message from which the items can be parsed out, using a syntactic analysis. Too complex for our simple Agents, it may be ultimately needed in high-level thinking, communication, and encoding of complex ideas in long-term memory [Note 3].

5. Send nothing! *The recipient already knows where to find its arguments. The recipient is activated by pattern-matching on the current state of the process.*

We are proposing a variation of the latter -- no message at all. We don't even need to notify the recipient. Our main purpose is to propose that:

Each of an agent's data sources is a FIXED location in (short term) memory.

Computationally, this means we are proposing to use "global variables", with all the convenience and dangerous side-effects well-known in computer programming. This idea is an extension of the "common terminal" idea in my paper on frame-

systems. [Minsky, 1974]

It is perfectly normal, in the outside world, to use fixed locations for fixed purposes. People do not repeatedly have to tell one another that one gets water from faucets, electricity from outlets, mail from mailboxes, and so forth. If such conventions are not rigidly followed, there will be misunderstandings. The developmental proposals in this paper explain why that should not be a problem, at least in the early stages.

Short Term Memory: Although contemporary mind-theories seem to agree that there is a central "short term memory", "STM" for short, I have not seen discussed much whether specific elements of STM have specific functions. We are not adopting the standard STM theory in which a small number of common units are shared by all processes. We suggest that STM really is an extensive, branching structure, whose parts are not interchangeable; each has specific significance for the agents attached to it. *In fact, the agents are the STM.*

There are well-known experiments in cognitive psychology that are usually interpreted to show that there are a limited number of STM units. We interpret them as showing, instead, that different groups of agents block one another so far as external communication is involved. In any given experiment one will get "just so far" into the memory-tree before fluent communication breaks down. Different contexts expose different fragments of this tree, so the totality of STM is really very extensive.

Postulating inflexible, specific memory-connections raises serious problems. We will suppose, for example, that the "instrument" of a proposed action would usually be specified by a particular STM unit. Where does this specificity end? Is there a fixed assignment for, say, *the color of the instrument*, as in "break the glass with the green hammer?" The answer would depend on many factors, especially upon how important is each particular concept to each person. But there must be some end to ad hoc structure and eventually each intelligent person must develop a systematic way to deal with unfamiliar descriptions. We shall return to this later, in proposing a "case-shift" mechanism.

How could there possibly be enough STM "locations" to serve all such purposes? The restriction may not be so severe as it might appear, if we think of these memories as analogous to the limited variety of "cases" in a natural language. Nothing prevents an agent from treating one of its arguments in an unusual way, but to do this it must find a way to exploit the

> conventions for its purposes. We get by, in natural language, by using context to transcend formal surface limitations -- e.g., as when different verbs use the same prepositions in different ways. [Note 4]

These problems will lead us, further on, to consider the more general problem of focussing attention to subsidiary functions and subgoals. Because this issue must be treated differently for infants, children, and adults, the next section discusses the methodology of dealing with such problems.

Methodology: Performance vs. Development: In some ways this paper might appear a model of scientific irresponsibility, with so many speculations about which there is so little evidence. Some workers in Artificial Intelligence may be disconcerted by the "high level" of discussion in this paper, and cry out for more lower-level details. At this point in our thinking most of such detail would be arbitrary at best, and often misleading. But this is not only a matter of default. There are many real questions about overall organization of the mind that are not just problems of implementation detail. The detail of an AI theory (or one from Psychology or from Linguistics) will miss the point, if machines that use it can't be made to think. Particularly in regard to ideas about the brain, there is at present a poverty of sophisticated conceptions, and the theory below is offered to encourage others to think about that problem.

Minds are complex, intricate systems that evolve through elaborate developmental processes. To describe one, even at a single moment of that history, must be very difficult. On the surface, one might suppose it even harder to describe its whole developmental history. Shouldn't we content ourselves with trying to describe just the "final performance?" We think just the contrary. Only a good theory of the principles of the mind's development can yield a manageable theory of how it finally comes to work.

In any case, the next few sections outline a model with limited performance power, to serve in early stages of intellectual development. Adults do not work the same way as infants, nor are their processes so uniform. While adults appear on the surface to construct and explore GPS-like recursively constructed goal-trees [Note 7], we need not provide for such functions in infants. We must, however, explain how they could eventually develop. Similarly, in linguistic matters, we should not assume that very young children handle nested or embedded structures in their perceptual, problem-solving, or grammatical machinery.

We must be particularly cautious about such questions as "what sorts of data structures does Memory use?" There is no single answer: different mechanisms

succeed one another, some persist, some are abandoned or modified. When an adult recognizes (albeit unconsciously) that he needs to acquire a substantial new skill, he may engage in a deliberate, rather formal process of planning and problem-solving, in which he designs and constructs substantially new procedures and data-structures. To explain that sort of individualistic self-shaping, the proper form of a "theory of mind" cannot focus only on end-products; it must describe principles through which:

An earlier stage directs the construction of a later stage.
Two stages can operate compatibly during transtion.
The construction skill itself can evolve and be passed on.

Genesis of Fixed Location Assignments: How might Agents come to agree about which memory units should contain which arguments? The problem seems almost to disappear if we assume that

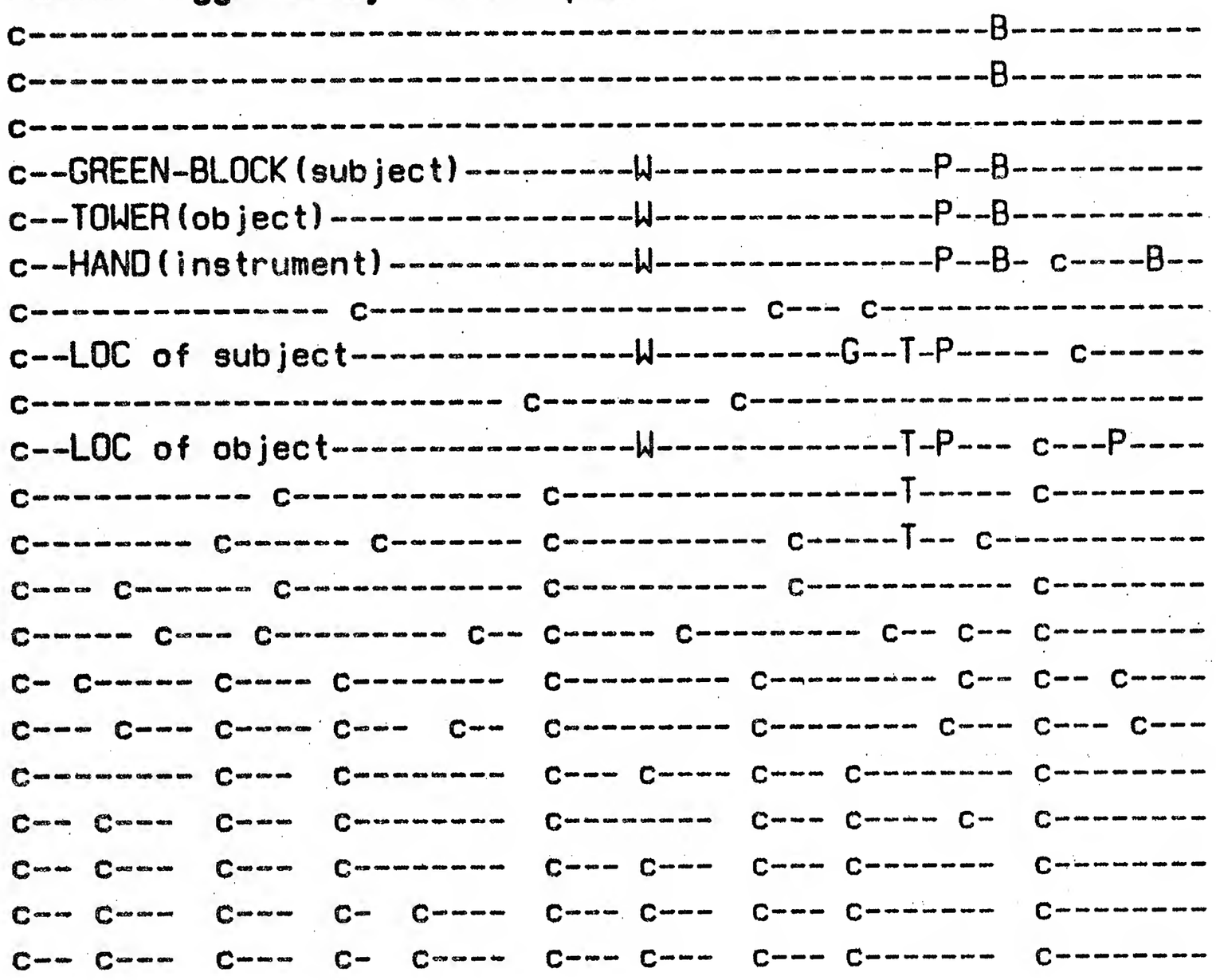
The agents evolve while continuously under the fixed-location constraint.
New agents arise by splitting off from old ones, with only small changes.
Thus they are born with essentially the same data connections.

This is appropriate because as new agents emerge, we expect them mainly to serve functionally as variants of their ancestors. To be sure, this does not account for introduction of radically novel agents but, just as in Organic Evolution, it is not necessary to suppose that this happens often -- or even ever -- in the infancy of an individual personality. And we are not so constrained in later life, for with the advent of abstract plans, and higher-level programming techniques, one can accomplish *anything* in a pre-planned series of small steps.

None of this is meant to suggest that all early agents are similar to one another; quite the contrary. Agents in different parts of the early brain surely use a variety of different representation and memory structures. In fact, we would expect distantly-related families to use physically separate memory systems. The price: agents concerned with very different jobs will not be able to communicate directly across their "social boundaries".

The figure below suggests an anatomical hierarchy of differentiation in which spatially different sub-societies emerge with connections through higher levels. At the top of this hierarchy are a very special few units -- of the very earliest genesis -- that serve to coordinate the largest divisions of the whole Society. These units, we speculate, lie at the root of many cognitive, linguistic, and other psychological phenomena.

Communication Cells and the "Specificity Gradient": We imagine the brain to contain a vast ensemble of "agents" connected by communication channels in a manner suggested by this example:



Each agent connects with a few near-by channels -- we'll call them c-lines. The diagram barely hints at the magnitude of the system - we see the c-lines as forming the vast network of the brain's white matter, with the agents forming the cortex. Descending, the structure divides and branches and, at lower levels, the agents become segregated into smaller and smaller sub-societies. They communicate within, but not between, those divisions. As in any highly developed society, effective communication with the outside, or between the largest subdivisions, usually must pass through the top levels. But see Note 9.

In the diagram a "high-level" agent B (for BUILDER) shares some terminals with another, P (for PUT), and also with yet another, W (for WRECKER). P shares some terminals with a lower-level agent T (for TRAJECTORY), which has none in common with B. Then B and W might be equivalent for some jobs (i.e., what should I do with this tower?) but not others. Deeper in the network, sub-sub-specialists can communicate directly only within localized communities. We will use the term *specificity gradient* for this gradual decentralization.

Each high-level channel must be a major resource, because it extends over a substantial portion of the brain and agents in many communities can interact through it. There cannot be very many of them, and their establishment is a major developmental influence.

Neurological Speculations: the "Laminar Hypothesis:" We identify this concept with the gross anatomy of the brain-- -- but with no pretense that there is any solid evidence for it. We suppose that many "innate" functions are genetically established by shaping the gross architecture of neural tracts -- great parallel, multiple bundles of pathways. We suppose that in infancy these functions are initially realized in these grossly redundant bundles, with the same computations duplicated in many, nearly parallel, layers. In the early stages these multiple systems act like single units (we speculate) because their components are functionally tied together by some form of "crosstalk" provided for this purpose. Later, these redundant layers -- we'll call them "laminae" -- slowly differentiate, as the crosstalk interaction is reduced under genetic control. [Note 5] Then agents in nearby laminae can begin to function independently as influenced by progressively more specific trigger conditions. This differentiation might proceed partly along the lines of Winston's learning paradigm -- in which clear, specific "differences" cause specific modifications within a differentiated agent -- and partly along the lines of a complementary process, "concept-leaf separation" -- in which agents within a family become competitive or mutually exclusive, each representing a different "sense" of the same "concept". [Note 6]

Both the communication paths and the attached masses of potential agents undergo the same sort of evolution together. At first, genetic control enables only large bundles to function as units. The community of agents of infancy would use these to build simple, basic, representations of objects, actions, obstacles, goals, and other essential symbolic entities. Early cognitive life would center around them. Once reliable uses of "top-level" agencies are established, we can begin to differentiate out families of variant sub-specialists that share local data-lines because of their common origins.

Members of different lower-level families, even at the same level, cannot communicate sideways -- for two reasons, functional and anatomical. Anatomically, we know that on a local scale the fibres of bundles tend to lie parallel, but on a larger scale they divide and branch. This pattern repeats over many orders of scale. Functionally, as we proceed further into specialization, the data-variables that concern processes also become more local. Sub-processes concerned with specialized sub-problems, or with different views of parts of a problem, need fewer common global symbols.

Genetics and Connections of agents: A typical agent or process (we suppose) uses c-lines on perhaps two or three adjacent levels. As jobs become more specific, their computations "descend" into progressively specialized brain regions. Cross-communications must be relayed up and down again through overlapping agencies.

The tree-structure of the C-diagram oversimplifies the situation in the brain. The divisions are not arbitrary and senseless, but under the most intricate genetic control. For example, if agents dealing with motions of the arm and hand are localized in one area, we would expect genetics to provide some common c-lines to an area concerned with visual recognition and scene analysis. After an early period of sensori-motor development these might be superceded by more "general" connections. If these general principles are on the right track, then the gross functional neuroanatomy should embody the basic principles of our early innate developmental predispositions, and finer details of the "genetic program" are expressed in the small details of how the regions are interconnected. It goes without saying that this is true also at the most microscopic levels, at which the genetic programs establish the different properties of agents in different, specialized, areas.

In infancy, connections to the top-level, common channels are made to simple production-like agents through relatively short path-chains. These learn to augment innate, specific, instinctual mechanisms by adding situation recognizers and motor patterns selected by internal motivational events; these become ingredients for later representations.

Overall coordination of the whole system needs an elaborate and reliable instinctual structure -- and we like the general ideas of the cross-linked hierarchical model proposed by Tinbergen [Note 5] for animal behavior. In that system, the different behaviors associated with different "basic" motivational drives employ substantially separate agencies, with coordination based on a priority-intensity selection scheme. Tinbergen's "modules" have an interesting provision for what is, in effect, heuristic search -- the "appetitive" connections in the hierarchy. Later, more coherent, knowledge-based ego-structures must emerge to supervise the system.

So we imagine the system beginning life, with a simplified skeleton of a later system (but not the final one). Each special area, with its mass of potential agents is connected internally and externally by gross bundles that first function as units. As development proceeds, the simple *sensory-->common-->motor* connections elaborate into the stratified, hierarchical structure pictured above.

Communication: How should agents read and write onto the communication lines;

what symbols should they use? From a point of view in which the agents are so simple that "meanings" are inaccessible, does it make sense to read or write anything at all? When this problem was first faced in the early work of Newell, Shaw, and Simon, it became clear that in a very low-level symbol-manipulation system one had to reduce the concept of "reading" to operations of matching and testing, not of "meaning".

Agents of any family related closely enough to share many terminals would tend to have common origins and related competences. They would usually constitute a "frame-system" or a "branch of concept-leaves", in which the choice of which branch or leaf gets control can often be made on the basis of local context. We suggested above that agents take inputs from several nearby levels. The highest of these could be seen as addresses, enabling groups of perhaps competitive agents. Middle levels could be seen as "context" for which of these has priority, and lowest levels as data. Agents whose outputs are above their inputs are useful in analytic recognition, e.g., parsing or scene-analysis. Agents with outputs below are "method" or "effector-like", activativating lower-level sub-processes.

How are connections learned? Probably symbols are represented as parallel patterns recognized by simple, perceptron-like detectors. These are attractive because of the simple training algorithms available; local perceptrons share some features with hash-coding. In particular, the surface representation can be meaningless. While perceptrons are not good at complex tasks, they seem appropriate for local symbol and coincidence learning jobs. [Note 8]

Temporary Storage, and Recursion: AI workers following the tradition of recursive pursuit of subgoals, often consider theories in which, *when an expert passes control (temporarily) to another, STM memory is pushed onto a stack and its contents reassigned.*

Given at least some access back to the push-down stack, this makes the full power of the intelligent machine available for the pursuit of subgoals. Adults ultimately find ways to focus "almost full attention" on subproblems, without completely losing track of the main problem [Note 7}.

In younger people, though, this is probably not the way; full attention to a subproblem may leave no road back. But, long before a person is able to put one problem entirely aside, work on another, and return to the first [Note 7], he must develop more limited schemes for withstanding brief interruptions. We will propose a mechanism for this; it has two components:

Persistence-memory: *The c-lines (or the agents driving them) have a*

tendency to restore recent sustained states -- that is, that they have a slower persistence-memory, so that when a transient disturbance is removed the preceding activity tends to be restored.

Transient Case-Shift: *We will also assume a mechanism for "shifting" patterns "upward" from at least some levels of the c-line hierarchy, on command from special agents that control these (later-maturing) pathways between the layers.*

The persistence memory means that if the child's attention is drawn to another subject, his present commitments are suppressed, for the moment, while a new Society is assembled. At the end of the diversion, interruption, or sub-job, the passive persistence memory restores the previous society -- which then must readjust to whatever changes have been made. For infantile problem-solving even just one such level would suffice for refocussing transiently on some kinds of subproblems.

Would data so crudely shifted remain meaningful? Only to the extent that adjacent levels have similar structure. This would be only approximate and many agents will get inappropriate messages. Our c-diagram does not illustrate this idea very well; there would have to be similar layers at each level. I should add that this case-shift idea seems physiologically unnatural to me. Its motivation and power will be seen below in "Minitheory 3," but something about it bothers me and it is proposed more as an exercise than as a strong conjecture.

General Memory: In later life we need larger, more "general purpose" memories. As the system matures, new families of memory agents could become actively attached to the c-lines. We need a concept of how they are addressed, and the issues there seem very similar to those of communication. It would seem plausible to have some sort of

Adjacent-context addressing: a memory unit is evoked by a pattern on one c-line level and remembers the contents of c-lines of an adjacent level. Recall activates, competitively, the memory agent that best matches the current address pattern.

Use of the *upper* level as the address makes the system suitable for activating sub-agents; use of the *lower* level as address is useful for reporting, e.g., for agents that recognize complex things by relations between their parts. There is a deep question about whether the same knowledge-bearing agents can be used in both ways -- as "antecedent" and as "consequent" directed, to use Hewitt's distinction. Do we use the identical grammar agents both for talking and for

listening? More generally, do we share the same agents for explaining and for predicting? Who knows. In [Minsky, 1974], I suggested a two-way process in which "frames" would try to connect up by matching at both levels. I still don't understand the issues very well.

In any case, in this arrangement the memory units have the same kind of "split-level" connections as do other agents. Are they different from other kinds of agents at all? Probably one could construct a "unified theory". But, in the brain, such economy seems inappropriate: surely memory is important enough for a few genes of its own! Perhaps memory agents are even complex enough to sense differences and make simple changes of the Winston sort, within simple semantic networks.

It seems useless to propose too many details at this level. For, once the system has facilities for long-term memory -- that is, for restoring some parts of the c-system to a semblance of an earlier state, the "mind" is capable, at least in principle, of constructing within itself new ways to learn, and to "think" about what it is doing. It can go on to build for itself more powerful and more general capabilities. At this point the mechanisms of higher thought become decoupled from basic computational principles, and from basic gross anatomical features, and the methodology of correlating structure with function begins to fail.

Emergence of cognitive "universals": Do all people think the same way? How is it possible for them to communicate? Are important features of Thought and Language determined in precise detail by genetics, or only through broad developmental principles? All natural languages, we are told, have much the same kinds of nouns, verbs, adjectives, cases, and even (it is said) some invariances of word-order. They certainly all have *words*. Is this because of a highly-structured, innate linguistic mechanism?

The question is important, but not so much for the theory of Syntax as for the theory of Thinking in general. One possibility is that detailed syntactic restrictions are genetically encoded, directly, into our brains; this raises problems about the connections with meaning and the representation of knowledge. Another possibility is that there are uniformities in early cognitive structure that affect the evolution of languages, both in the individual, and circularly, through the culture. In the social evolution of child-raising, cultures learn what is easy for children to learn; early language cannot say what young children cannot "mean", to use Halliday's expression. And much of what children "mean" develops before overt natural language, within the high-level internal c-lines of our theory. The rest of this paper pursues what this might imply about pre-linguistic meaning.

It would seem inevitable that some early high-level representations, developing in the first year or so, would be concerned with the "objects" of attention -- the "things" that natural languages later represent as nouns. Here we would indeed suspect genetic pre-structuring, within sensory systems designed to partition and aggregate their inputs into data for representing "things". Further, we would need systems with elementary operations for constructing, and comparing *descriptions*. We pursue this by returning to the action PUT: what is required of such an Agent? Setting aside possible conceptions far outside the present framework, let us agree that PUT needs access to c-lines for ORIGIN (green block) and DESTINATION (tower-top).

What, in turn, does it mean for there to be ORIGIN c-lines? In the introduction, we pointed out that different sub-agents of PUT will need to know different things about the ORIGIN; MOVE will need location-of-origin and GRASP will need size-of-origin. Consider a model in which the description is simply a property-list in which the value of a property is a symbol (in binary) on a bundle of c-lines.

MINITHEORY 1: *Somewhere there is an agent, G, with access to the property list of GREEN-BLOCK. The agent G knows -- or controls subordinates who can find out -- some things like color, size, shape and location. When we say that "GREEN BLOCK" is the value of ORIGIN, we mean that the activity pattern on the origin c-lines somehow activate this G to give it dominance over potential competitors at its own level.*

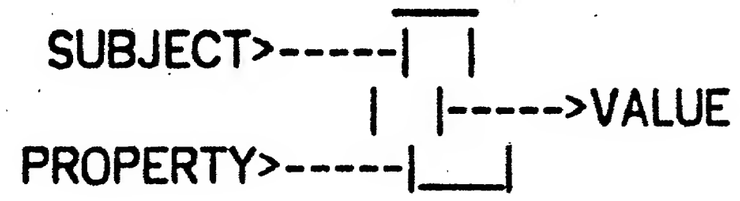
Activation of G, in this infantile minitheory, simply enables it (or its subordinates) to place property-value symbols onto certain other c-lines, e.g., on c-color-of-origin, c-size-of-origin, etc. This makes the description available to other agents like PUT. But is it reasonable to suppose a distinct c-line for every distinct (known) property of the subject?

For adults, this would seem extravagant. For an infant, it seems no problem -- he's lucky to have any properties at all. Later in development, though, there will be other "conceptual foci" -- let's call them conceptual cases in analogy with linguistic cases -- such as OBJECT, INSTRUMENT, INDIRECT-OBJECT, and so forth. Will we have to reduplicate the whole property structure all over again for each of these?

What might these "cases" be? One theory is that we begin with a single object or noun-like case which later splits into two or more, used to represent goals and effects of actions. The *actions* -- "verb-like" structures -- could remain implicit in the agents till later. We will consider other possibilities shortly; for the moment we will carry on as though there are just a few similar noun-case

structures. But if they are very similar, the proliferation could be stemmed by creating a more general kind of noun-agent to take care of such matters. We can imagine two somewhat opposite approaches to this.

MINITHEORY 2: *We create a NOUN-AGENT who functions as a general-purpose property-selector. It has just two input c-lines and one output c-line. (In fact, it is just the LISP function, GETPROP.)*



By providing just one of these for each functional noun-case, we are saved from duplicating each c-line for each property. Unfortunately, there are some serious objections to this scheme.

> Having only one property available at a times makes difficulties for "recognition agents". As in "production system" or "PLANNER-like" models, our agents do not usually specifically designate their sub-agents; rather, they "enable" whole sub-families, whose members select one of themselves to respond on the basis of other data, e.g, recognition of important combinations of properties. In serial computers this would be no problem, but here the system would have to scan through all the properties.
>
> A more subtle but more serious problem: the specificity of the c-lines has been lost. The same c-line sometimes means size, sometimes location, and so forth. The poor agents will find most of their inputs totally meaningless. We have lost "homogeneity of symbol-type".

MINITHEORY 3: *Let's agree instead to tolerate Minitheory 1's set of distinct c-lines for each property -- these would be indispensible in infancy. Then the recognition agents could be simple, one-layer perceptrons. [Note 8] But, we don't want to duplicate all the agents, too, for each case. Imagine that the several cases all differentiate from a one primordial case, as a linear sequence with specificity gradient:*

```
NOUN1<---NOUN2<---NOUN3<---NOUN4<---
 kind      kind      kind      kind
 loc       loc       loc       loc
 size      size      size      ---
 origin    origin     ---      ---
 shape     shape      ---      ---
 destination --     destination
 support   support    ---      ---
 purpose     ---      ---      ---
 color       ---      ---      ---
 etc.
```

A huge advantage of this is that agents can have access to several objects at once, so that they can recognize differences and other relations between objects. This makes it possible to learn simple forms of "means-ends" analysis, etc.

The most highly developed case has the best-developed description structure while the others have progressively smaller (and somewhat different, more specialized) To embody the "case-shifter" descussed earlier we can

> arrange the c-lines for the different case-symbols so that a case-shift mechanism can move their contents into better-developed case-slots.

Now we can move, transiently or permanently, any noun-symbol chosen as the focus of attention, into a more principal case-position. *Then a more detailed description of the selected object of attention appears on the property c-lines of the "more principal" case.* There are substantial advantages to this arrangement:

> It preserves homogeneity of type. Each c-line always carries the same "sort" of information, so that the problems for sub-agent recognition are vastly simplified.
>
> The different case-systems provide different descriptions of the same object.
>
> The stratification of the case structure is very plausible, developmentally. The infant conception evolves from a single object focus -- neither SUBJECT or OBJECT, but just IT.

The linear layout of the diagram suggests that the whole sequence might be shifted at once. If this were true, it suggests a prediction that there might be a

preferred ordering of cases in natural language when it later appears. Suppose that OBJECT (assuming, arbitrarily, it to be the "second" case) is shifted into SUBJECT, the "first" case. Then some certain "third" case will, by default, usually shift to replace the OBJECT. Is there a linguistic regularity, in early language development, anything like this? If so, it might reflect a remnant of this primordial ordering.

Are the Cognitive Cases "Universal?" We just passed over a vital question, in assuming that the functional cases are the same from one person, or culture, to the next! The original IT represents the central object of attention in the infant. When IT later splits into two, these might be used for finding differences (as suggested above), or they might be involved in describing things as wholes made of parts -- the beginnings of true description. Another possibility is that they are first involved in *before-after* or *have-want* representations of actions and of goals. Is there a distinct, early, *agent* case -- and how is it involved with the infant's representation of himself as different from the rest of the world? And what comes next? Is it an *active-agent - passive patient* distinction, an *object-action* distinction, an *instrument-object* distinction, or what? It seems to me that this question of whether there are genetic or computational reasons for one structure or another to first appear should be a central issue in the theory of human intelligence. Perhaps the study of earliest language features can help here.

Minitheory 3 still leaves serious problems, if we want to probe more deeply into descriptions. For example, the concept "INSIDE", say, of a box, should yield another object -- not merely a property. Is INSIDE a property? Obviously, the idea of description as property-list itself has limitations, and we cannot simply continue forever to add more and more properties, with dedicated c-lines. Concepts like contents-of, opposite-of, supported-by, or a-little-to-the-left-of presumably involve relations between cognitive cases. And what about scenes in which there are rows of rows, arches of arches, etc. As our descriptive power increases, so must that of the agencies employing the descriptions, and simple, uniform solutions like the case-shift mechanism will not suffice.

I see little use in trying to attack such problems by further naive psychophysiological speculation. This is a job for Artificial Intelligence! One approach is to find ways such systems could uniform and universal solutions. Thus, one might look for ways to make some agents to embody the primitive operations of LISP, while others learn to embody representations of LISP programs. Or, perhaps, one might apply a similar plan to some adequate logical formalism.

Another approach is to search "basic" or "primitive" operations that, while

perhaps less elegant and uniform, seem more lifelike. For example, mental activity surely needs processes that can:

Compare two descriptions.
Describe the result of a proposed action.
Find an action that will reduce a difference.

Yet another approach is to search for a coherent basis of "conceptual" relationships, or "dependencies", as Schank has put it. Here one might focus on the representation issue first, and hope that the procedural issues will clarify themselves.

At some point in each of these plans, the strategy should turn toward seeking some developmental uniformity. I would expect the "final, true" scheme to turn out to seem wildly *ad hoc* on the surface. Imagine trying to account for the stick-insect, without understanding its evolutionary origin.

Description and Language: About the simplest kind of verbal description is that of noun plus adjective. Mini-theory 1 can represent any particular such description, e.g., *size-of-subject*, in an ad hoc fashion, by creating a specific c-line for it. Through a connection to such a c-line a child could learn, "by rote" or whatever, to respond to such an object with a "pseudo-syntactic" verbal form like "Large Block". The syntax would be "pseudo" for lack of a systematic way to construct -- or understand -- other such forms. A truly syntactic development would be the ability to attach any available adjective to any noun-case object. With Minitheory 3, on the other hand, one could imagine fragments of true syntax emerging after completion of fragments of the attention-focus case-shifter, for this would provide at least the rudiments of appropriately meaningful deep structure.

But I don't think it plausible to expect complex verbal behavior to follow so closely on the heels of barely completed cognitive machinery. For one thing, it isn't true in real life; the development of internal representations seems much further ahead in the first two years. And we'll propose a hint of a possible theoretical problem in a moment.

In any event, more elaborate "actions", "frames" and "scenarios" surely become mental objects after the first few months. We wish we knew a way to tell whether these in fact take the form of cross-cultural "cognitive universals", in the sense that there are comparatively similar representations between one child and another. There seems no question that elements which seem necessary to summarize the simplest real-life episodes bear compelling likenesses to familiar

linguistic structures. Thus, specification of the *instrument of the action* ("by" or "with"), the *purpose* ("for"), the *trajectory* if provided ("from -- to") and so forth, seem essential. It can hardly be a coincidence that these entities, which appear later in language, resemble so the ingredients that seem earlier needed for what we might call "cognitive syntax". Or -- gloomy possibility -- perhaps this is just an illusion that stems from cognitive contamination by the structure of one's own natural language! [Note 7]

When the time finally comes for learning grammatical speech, those "cognitive cases" that have high-level c-representations should be relatively easy to encode into external symbols, because the "deep structures" (along with some means for manipulating them) already exist as central elements of internal communication. That is, the child already uses something like syntactic structure in his internal manipulations of descriptions. In fact, let us reverse the usual question about how early children learn to talk grammatically. Perhaps we need a theory of *why do children take so long to learn to talk grammatically!*

The phenomenal suddenness with which many a child acquires substantial grammatical ability, around his second year, certainly cries out for explanation. But why not sooner if, internally, he already uses something as complex? Conjecture: it is a matter of a different kind of computational complexity -- of converting from one sort of data-structure to another. For, if we admit an earlier "non-natural internal language" of a sort, then *learning language is really learning to translate between languages!*

Well, what could cause a sudden growth in computational power? Conjecture: it is not until somewhere in his second year that his computational development takes some important additional step along the road to the "full" computation power that makes "almost anything" possible. We know a great deal of theory about such matters, not in cognition, but in the theory of computation. And there we are used to seeing how very simple changes can make large surface differences -- e.g., in adding another level of approximation to recursion. So the answer to our question might lie partly in that other arena, wherein the addition of an inconspicuous new computational facility makes a dramatic (but now non-mysterious) increase in symbol-manipulative competence. In any case, the discovery and study of "linguistic universals" promises to provide us with deep and important suggestions about the structure of internal communication.

This leads, I think, to the following position about "innate vs. acquired" in language: The internal communication mechanisms in the infant mind, at least at the higher levels, may have enough uniformities to compel society, in subtle ways, to certain conformities, if social communication is to engage young children. While

the fine details of mature linguistic syntax could be substantially arbitrary (because they develop comparatively late and can exploit more powerful computational mechanisms), they probably are not so, because the cognitive entities that early language is concerned with are probably much more rigidly and uniformly defined in infancy.

NOTES

NOTE 1. The present paper is in part a sequel to my paper [Minsky 74] and partly some speculations about the brain that depend on a theory being pursued in collaboration with Seymour Papert. In that theory, which we call "The Society of Minds", Papert and I try to combine methods from developmental, dynamic, and cognitive psychological theories with ideas from AI and computational theories. Freud and Piaget play important roles. In this theory, mental abilities, both "intellectual" and "affective" (and we ultimately reject the distinction) emerge from interactions between "agents" organized into local, quasi-political hierarchies. Overall coherency of personality finally emerges, not from any clear and simple cybernetic principles, but from the interactions, under elaborate genetic control, of communities of do-ers, "critics" and "censors", culminating in almost Freudian agencies for self-discipline that compare one's behavior with fragments of "self-images" acquired at earlier stages of development. The PLAY episode that begins the present paper hints briefly at the workings of such an internal society under the surface of a child's behavior. We hope to publish the whole theory within the next year or so, but it still has rough spots.

NOTE 2. Because (1) Wrecker was the primary goal and (2) his job is easier than BUILDERS's in the limited time-frame and (3) it satisfies some more remote goal which is angry at HUNGER's subversion of PLAY and (4) the consummatory act that closes the episode leaves less unfinished business and conflict.

NOTE 3. We do not mean to preclude the use of natural language for internal purposes. Everyone agrees that it happens all the time. But these are not conversations between simple agents, but emerge from interactions among vast, organized societies. A developed personality is an enormous structure that has constructed for itself fantastic facilities, in which some parts can send verbal messages, and other, structured representations, to other parts of itself -- .or even to future, contingent activations of itself.

NOTE 4. It is natural to wonder why we tolerate so much ambiguity as we do in natural language. Conjecture: we hardly notice it because we have developed such powerful methods for dealing with ambiguity in other mental forms. The forthcoming work with Papert will propose, as a main thesis, that thoughts

themselves are ambiguous in an important sense. Then the "disambiguation" of natural-language expressions is "child's play" compared to other conceptual representation problems.

NOTE 5. We conjecture that there is a general mechanism through which the neurological structures responsible for learning pass from early stages in which collections of cells act as units to later stages in which their components become mutually cross-inhibitory. That is, they move toward an "exclusive-or" mode of operation in which only one component at a time can get control of a group's shared output ports. Such families might resemble the form described in the "synthesis" diagram in Chap.V of Tinbergen's A Study of Instinct; both the agents for "real-time" and for for long term memory might derive from modifications of such structures.

NOTE 6. A main theme of the work mentioned in Note 1 is the idea that "learning" often involves a choice of whether to try to modify an old concept to a new purpose, or to split it into two variants. Winston's "near miss" technique tries to make a single representation accomodate to new problems. When this doesn't work, we conjecture, the representation splits into two or more "leaves" of a competitive family.

NOTE 7. The "recursive" use of mental facilities, easily "simulated" with programming languages like LISP is probably an illusion, an artifact of our description of what we observe ourselves doing. A recursive function call is just one, extremely clean and simple way to separate the local variables from one "invocation" of a process to another. The infantile schemes proposed in this paper lie at another extreme. Presumably, as people mature they construct other intermediate forms with various features and bugs. A problem is that we don't have enough technical names of theories for other "approximate" schemes for "context maintenance". (Comment suggested by G. J. Sussman.) See for example, [McDermott & Sussman]

NOTE 8. The formal limitations of simple perceptrons should not be troublesome, here, because we can assume that each input c-line carries a comparatively meaningful information symbol from some other well-developed system. In such circumstances, the perceptron learning algorithm, "on tap, not on top", could have immense value since, conceivably, it could be embodied in just one or a few brain cells.

NOTE 9. No moral is intended in this analogy. Just as in human societies, there are surely important ways in which low-level agents cross gross boundaries, and these may have vital neurological functions. However, the analogy is poor

because human individuals can know and understand more than their social superiors; this can hardly happen in the nervous system, in which all the agents are equally mindless.

REFERENCES

Halliday, M.A.K., *Learning How to Mean* - Explorations in the Development of Language, Edward Arnold (Publishers) Ltd., London, 1975

Hewitt, Carl. E., *PLANNER: A Language for Proving Theorems and Manipulating Models in A Robot*, MIT Artificial Intelligence Laboratory Technical Rept. 258, 1972

McDermott, Drew and Sussman, G., *The CONNIVER Reference Manual*, A.I. Memo # 259A, May 1974.

Minsky, Marvin, *A Framework for Representing Knowledge*, MIT Artificial Intelligence Laboratory Memo. No. 306, June 1974. Also in *Psychology of Computer Vision*, (P.H. Winston, Ed.) McGraw-Hill, 1975.

Newell, A. and Simon, H. A., *Human Problem Solving*, Prentice Hall, 1972.

Tinbergen, T., *The Study of Instinct*, Oxford University Press, 1951.

Winograd, T. *Understanding Natural Language* Academic Press, N.Y., 1973

Winston, P. H., "Learning Structural Descriptions by Examples", in *Psychology of Computer Vision*, (P.H. Winston, Ed.) McGraw-Hill, 1975.